View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace at VU

TI 2001-110/3
Tinbergen Institute Discussion Paper

Rounding of Arrival and Departure Times in Travel Surveys

Piet Rietveld

Department of Spatial Economics, Faculty of Economics and Business Programming, Vrije Universiteit Amsterdam

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl

Rounding of Arrival and Departure Times in Travel Surveys;
an interpretation in terms of scheduled activities.

2001

Piet Rietveld
Vrije Universiteit
Amsterdam
The Netherlands
prietveld@econ.vu.nl

Abstract.

In travel surveys most respondents apply rounding of departure and arrival times to multiples of 5, 15 and 30 minutes: in the annual Dutch travel survey about 85-95 percent of all reported times are 'round' ones. We estimate rounding models for departure and arrival times. The model allows one to compute the probability that a reported arrival time m (say m=9:15 am) means that the actual arrival time equals n (say m=9:21 am). Departure times appear to be rounded much more frequently than arrival times. An interpretation for this result is offered by distinguishing between scheduled and non-scheduled activities, and by addressing the role of transitory activities.
We argue that explicitly addressing rounding of arrival and departure times will have at least three positive effects. 1. It leads to a considerably better treatment of variances of reported travel times. 2. It enables one to avoid biases in the computation of average transport times based on travel surveys. 3. It overcomes the problem that the use of travel survey data for the minute-per-minute records of the development of the number of persons in traffic displays erratic patterns.

Keywords, travel time, transitory activities, rounding, Bayesian approach, scheduling

1. Introduction.

Research on travel behaviour is often based on travel times and distances reported by travellers. It is well known that these reported values tend to be rather inaccurate. This can be understood for distances travelled since many travellers do not have instruments to measure distance. In the case of travel time one might expect a more accurate measurement since most travellers have their own watch; in addition, travellers have to keep a close look at their travel times when they want to arrive in time for scheduled activities. Nevertheless, it is clear that also here inaccuracies occur (see for example Rietveld et al., 1999). Some people take clock time more serious than others, and there are also notable differences between cultures in the precision of timing activities (Levine, 1997). In the present paper we address the issue of rounding of travel times, more in particular the rounding of arrival and departure times.

Consider the example of reported departure times of trips in the annual national transport survey in The Netherlands (CBS, 1998). This survey is based on the travel diaries of about 144,000 randomly drawn Dutch citizens who report their travel activities during one day in the year 1997. Respondents are requested to report the arrival and departure times of all trips on a certain day. Suppose a respondent j indicates that a trip started at departure time $[h_j:m_j]$, where h_j indicates hour (h_j =0,1,...,23) and m_j indicates minute (m_j =0,1,...,59). Let q(m) denote the total number of respondents who reported to have departed at the reported minute m. Then Figure 1 contains the observed distribution of the minute of departure m of all respondents (m=0,1,...,59), where the hour h of departure has been deleted. The total number of reported departure times is 550,000 based on the questionnaires filled out by the 144,000 respondents. The figure shows extreme peaks in the distribution of reported departure times. It appears that about 22 per cent of all travellers report that they left at h o'clock sharp, (h=0,1,...,23), whereas this figure is only 0.14 per cent for travellers who report that they left at 1 minute past h o'clock. Multiples of 5 and 15 minutes also get very high shares. The share of reported departure times of non-multiples of 5 minutes is only 5 per cent, whereas their share on the clock is as high as 80 per cent (48/60). A similar pattern of reported departure times is observed in the US Nationwide Personal Transportation Survey (see for example Battelle, 1997).

Figure 1. Distribution of reported departure times.

For an adequate analysis of travel statistics it is important to pay attention to rounding because otherwise it would lead to unreliable data on travel times. For example, when somebody is used to round his departure and arrival times to multiples of 15 minutes, travel time itself will also be rounded to multiples of 15, implying a rather inaccurate reported travel time. This means that analysis of travel behaviour is based on inaccurate travel time data. A similar conclusion holds for the analysis of travel time budgets (see for example Zahavi, 1977) and travel speeds. The rounding problem adds another error to the usual errors in statistical analysis (incomplete data, specification error, fundamental unpredictability of human behaviour) and thus leads to larger variances of estimated coefficients.

Rounding does not only affect variances, it may even lead to a systematic bias for *average* figures. As we will demonstrate later on in the paper, there is no guarantee that in the case of travel times the probabilities of rounding upward and rounding downward are equal. Thus, rounding does not only affect the reliability of individual observations, but it may also have an adverse effect on the reliability of national averages. We will demonstrate that rounding practices provide an explanation of the result reported by Battelle (1997) that the average of reported travel times is higher than the average of actual travel times.

Another example of the problem with rounding is found when departure and arrival time data are used to describe the development of traffic volumes during peak periods. Travel survey data of the type discussed here can be used to find how many cars there are on the Dutch roads from minute to minute (see for example CBS, 1996), but rounding will lead to rather erratic patterns[1]. The simplest way to overcome this would be to present data during time units of 30 or 60 minutes, but this would imply that information is lost on how traffic volumes build up during the shoulders of the peak. This is important information for public and private actors that try to address congestion problems.

The above examples demonstrate the importance of rounding of departure and arrival times for data quality that affects transport analysis and policy making. However, the relevance of the topic of rounding of departure and arrival times goes beyond data reliability. We will demonstrate that the rounding phenomenon sheds light on the nature of scheduling of transport-inducing activities. We develop a simple statistical model to analyse the propensity to round departure and arrival times and estimate it in section 2. An interpretation for differences between rounding in departure and arrival times is given is section 3. We will do this by discussing the rounding phenomenon in the context of scheduled activities. Section 4 concludes.

2. Formulation and estimation of statistical model.

Formulation of statistical model

As Figure 1 shows, rounding of departure times seems to take place towards certain anchor points such as:

- multiples of 5 minutes: 0, 5, 10, 15, 20,...,55
- multiples of 15 minutes: 0, 15, 30, 45
- multiples of 30 minutes: 0, 30
- multiples of 60 minutes: 0

Note that according to this approach the high outcome for the [h:00] o'clock departure time in Figure 1 is the joint result of rounding to all multiples of 5, 15, 30 and 60 minutes. A fifth possibility is that people do not apply rounding but report the exact minute of departure.

Consider in more detail the possibility of rounding to the nearest multiple of 5. Let m be the actual minute of departure, and let d_{m5} be the absolute time distance to the nearest multiple of 5 (d_{m5}=1,2). For example, when m=23, the nearest multiple of 5 is 25 so that d_{m5}=2. Note also that $d_{59,5}$=1, since

[1] Note that if rounding were to take place to the same extent in both departure and arrival times the stock of people in traffic would not display erratic moves from minute to minute. However, when rounding is more prominent in one of the two processes, irregular patterns will be found in the minute to minute records of the stock of persons in traffic.

[(h+1):00] is the nearest multiple of 5 for [h:59]. The probability p_{m5} that the actual departure time m will be rounded to the nearest multiple of 5 is assumed to be[2]:

$$p_{m5} = a_5 + b_5.d_{m5} \qquad d_{m5}=1,2$$

The coefficient a_5 is interpreted as a base value for rounding to a multiple of 5 minutes, whereas b_5 indicates the decrease of the probability of rounding as one moves away from a multiple of 5 minutes. We expect a_5 to be positive and b_5 to be negative: there is a tendency of rounding towards the nearest multiple of 5 minutes, but this tendency decreases as one moves away from the 5-ple. For example, the probability of rounding 11 to 10 is larger than the probability of rounding 12 to 10. Note also that as p_{m5} has to be positive one must ensure that $a_5 + 2.b_5$ must be positive.

In a similar way we formulate the rounding mechanisms for the other multiples of minutes:

$$p_{m,15} = a_{15} + b_{15}.d_{m,15} \qquad d_{m,15}=1,2,..,7$$

$$p_{m,30} = a_{30} + b_{30}.d_{m,30} \qquad d_{m,30}=1,2,..,15$$

$$p_{m,60} = a_{60} + b_{60}.d_{m,60} \qquad d_{m,60}=1,2,..,30$$

In the case of rounding to a multiple of 30 minutes there are two nearest multiples when m=15. In this case the probabilities of rounding to [h:00] and [h:30] are assumed to be equal, so that the resulting probabilities of rounding are $(a_{30} + 15.b_{30})/2$. A similar case holds for the rounding to a multiple of 60 minutes.

After having defined these rounding probabilities, the probability that rounding of departure time m does *not* take place ($p_{m,0}$) equals:

$$p_{m,0} = 1-p_{m,5}-p_{m,15}-p_{m,30}-p_{m,60} \qquad \text{for all m, not being multiples of 5}$$

$$p_{m,0} = 1-p_{m,15}-p_{m,30}-p_{m,60} \qquad m = 5, 10, 20, 25, 35, 40, 50, 55$$

$$p_{m,0} = 1-p_{m,30}-p_{m,60} \qquad m = 15, 45$$

$$p_{m,0} = 1-p_{m,60} \qquad m = 30$$

$$p_{m,0} = 1 \qquad m = 0$$

Thus there is only one case where we assume that rounding does not take place, i.e., when m=0. The resulting structure of transition probabilities can be found in Table 1.

[2] Thus, p_{m5} can be interpreted as the *conditional* probability that given the actual departure time m the reported departure time is that multiple of five nearest to m.

	Time of departure in minutes reported by respondent given his actual departure time m								
Actual time of departure: m	**5-ple below m**	**5-ple above m**	**15-ple below m**	**15-ple above m**	**30-ple below m**	**30-ple above m**	**60-ple below m**	**60-ple above m**	**m (no rounding)**
0	0	0	0	0	0	0	0	0	1
1	$p_{1,5}$	0	$p_{1,15}$	0	$p_{1,30}$	0	$p_{1,60}$	0	$1-p_{1,5}-p_{1,15}-p_{1,30}-p_{1,60}$
2	$p_{2,5}$	0	$p_{2,15}$	0	$p_{2,30}$	0	$p_{2,60}$	0	$1-p_{2,5}-p_{2,15}-p_{2,30}-p_{2,60}$
3	0	$p_{3,5}$	$p_{3,15}$	0	$p_{3,30}$	0	$p_{3,60}$	0	$1-p_{3,5}-p_{3,15}-p_{3,30}-p_{3,60}$
4	0	$p_{4,5}$	$p_{4,15}$	0	$p_{4,30}$	0	$p_{4,60}$	0	$1-p_{4,5}-p_{4,15}-p_{4,30}-p_{4,60}$
5	0	0	$p_{5,15}$	0	$p_{5,30}$	0	$p_{5,60}$	0	$1-p_{5,15}-p_{5,30}-p_{5,60}$
6	$p_{6,5}$	0	$p_{6,15}$	0	$p_{6,30}$	0	$p_{6,60}$	0	$1-p_{6,5}-p_{6,15}-p_{6,30}-p_{6,60}$
7	$p_{7,5}$	0	$p_{7,15}$	0	$p_{7,30}$	0	$p_{7,60}$	0	$1-p_{7,5}-p_{7,15}-p_{7,30}-p_{7,60}$
8	0	$p_{8,5}$	0	$p_{8,15}$	$p_{8,30}$	0	$p_{8,60}$	0	$1-p_{8,5}-p_{8,15}-p_{8,30}-p_{8,60}$
9	0	$p_{9,5}$	0	$p_{9,15}$	$p_{9,30}$	0	$p_{9,60}$	0	$1-p_{9,5}-p_{9,15}-p_{9,30}-p_{9,60}$
10	0	0	0	$p_{10,15}$	$p_{10,30}$	0	$p_{10,60}$	0	$1--p_{10,15}-p_{10,30}-p_{10,60}$
11	$p_{11,5}$	0	0	$p_{11,15}$	$p_{11,30}$	0	$p_{11,60}$	0	$1-p_{11,5}-p_{11,15}-p_{11,30}-p_{11,60}$
12	$p_{12,5}$	0	0	$p_{12,15}$	$p_{12,30}$	0	$p_{12,60}$	0	$1-p_{12,5}-p_{12,15}-p_{12,30}-p_{12,60}$
13	0	$p_{13,5}$	0	$p_{13,15}$	$p_{13.30}$	0	$p_{13,60}$	0	$1-p_{13,5}-p_{13,15}-p_{13,30}-p_{13,60}$
14	0	$p_{14,5}$	0	$p_{14,15}$	$p_{14,30}$	0	$p_{14,60}$	0	$1-p_{14,5}-p_{14,15}-p_{14,30}-p_{14,60}$
15	0	0	0	0	$½p_{15,30}$	$½p_{15,30}$	$p_{15,60}$	0	$1-p_{15,30}-p_{15,60}$
16	$p_{16,5}$	0	$p_{16,15}$	0	0	$p_{16,30}$	$p_{16,60}$	0	$1-p_{16,5}- p_{16,15}-p_{16,30}-p_{16,60}$
.	.	.	.	.	.	.	.	.	.
29	0	$p_{29,5}$	0	$p_{29,15}$	0	$p_{29,30}$	$p_{29,60}$	0	$1-p_{29,5}- p_{29,15}-p_{29,30}-p_{29,60}$
30	0	0	0	0	0	0	$½p_{30,60}$	$½p_{30,60}$	$1-p_{30,60}$
31	$p_{31,5}$	0	$p_{31,15}$	0	$p_{31,30}$	0	0	$p_{31,60}$	$1-p_{31,5}- p_{31,15}-p_{31,30}-p_{31,60}$
.	.	.	.	.	.	.	.	.	.
59	0	$p_{59,5}$	0	$p_{59,15}$	0	$p_{59,30}$	0	$p_{59,60}$	$1-p_{59,5}- p_{59,15}-p_{59,30}-p_{59,60}$
60	0	0	0	0	0	0	0	0	1

Example: when the actual time of departure m is 8:16, rounding can take place to 8:15 (via $p_{16,5}$; nearest multiple of 5) , another time to 8:15 (via $p_{16,15}$; nearest multiple of 15), to 8:30 (via $p_{16,30}$; nearest multiple of 30) and to 8:00 (via $p_{16,60}$; nearest multiple of 60); the other possibility is that the actual and reported time of departure coincide (last column of table).

Table 1. Probability of rounding the actual time of departure m by a respondent to nearest multiple of 5, 15, 30 or 60 minutes (below or above m), or to m itself

Consider now the distribution of actual departure times. Let g_m denote the probability that a trip made by the respondent actually starts at minute m. Then, given the conditional probabilities of rounding formulated in Table 1 the *joint* probability of an actual departure time m and the reported value being its closest multiple of 5 is $g_m.p_{m,5}$. Thus, we can derive the resulting probability that departures are reported to take place at time m. For example, the table demonstrates that the probability of a reported time of departure of [h:45], denoted as q_{45} is the sum of probabilities of actual departures ranging from 38 to 52 minutes past h, each multiplied with its probability of rounding to 45 minutes:

$$q_{45} = [g_{38}.p_{38,15}+...+g_{52}.p_{52,15}] + [g_{43}.p_{43,5}+..+ g_{47}.p_{47,5}].$$

For the other departure times similar formulations can be derived. Note that for departure times m that are not equal to multiples of 5 we have simply:

$$q_m = g_m.[1-p_{m,5}-p_{m,15}-p_{m,30}-p_{m,60}]$$

We still have to formulate the distribution of actual departure times g_m. We will assume that *all departure times within an hour are equally probable*:

$$g_m= 1/60.$$

This assumption has to be made since we have no prior knowledge about the distribution of the exact minute in the hour during which departures take place[3]. Another assumption we make is that rounding is the only source of errors. Thus, we will not consider other sources of error, such as mistakes while filling out the survey questionnaire, inaccurate watches, etc. The possible implications of these assumptions are discussed at the end of the next section. These assumptions suffice for a specification of the likelihood q_m for all reported departure times m. Let N_m denote the actual number of times that departure minute m is reported by respondents. Then the resulting loglikelood of the reported departure time m is:

$$\ln L = N_0 \ln q_0 + N_1 \ln q_1 + ... + N_{59} \ln q_{59}$$

Under the null-hypothesis that reported departure times are equal to the actual departure times all probabilities in Table 1 are equal to 0, except the ones in the last column. This implies that

$$\ln L_0 = N_0 \ln(1/60) + N_1 \ln(1/60) + ... + N_{59} \ln(1/60) = N \ln(1/60)$$

where N equals the total number of observations.

[3] Of course we have rather accurate information about the distribution of departure times during the 24 hours of the day: during the night the number of departures is much smaller than at day-time. But very little is known about the distribution between the minutes within the hour.

Estimation of model; departure times.

The results of the maximum likelihood estimation for the departure minutes are reported in Table 2. The likelihood values indicate a strong support for the rejection of the null hypothesis. The test statistic $?^2 = 2(\ln L - \ln L_0)$ is asymptotically distributed chi-square with degrees of freedom equal to the number of restrictions on the parameters (8). The value of the test statistic corresponding to a 99 per cent probability of rejection of the null hypothesis is 20.1 in this case. We find an overwhelming evidence of the importance of rounding to multiples of 5, 15 and 30 minutes: their base values a_5, a_{15} and a_{30} are clearly significant. Only rounding to the 'whole hour' assumes a small value (a_{60} is smaller than 1 per cent). The b values are very small, with the exception of b_5: it indicates that the probability of rounding 4 to 5 equals 46.4 per cent, whereas for rounding 3 to 5 it equals 42.8 per cent. For rounding to multiples of 15, 30 and 60, the b values are positive, which is unexpected. Their levels are very small, however. The reason that some of them are significant is that the number of observations is large. Considering the magnitudes they assume, they can be ignored. Thus, we conclude that, with the exception of rounding to multiples of 5 minutes the rounding probabilities hardly depend on the distance to the reference value.

coefficient	Maximum likelihood estimate	standard error
a_5	0.500	0.00142
b_5	-0.0360	0.00075
a_{15}	0.284	0.00142
b_{15}	0.0016	0.00017
a_{30}	0.177	0.00149
b_{30}	0.00015	0.00008
a_{60}	0.0093	0.00108
b_{60}	0.00055	0.00004
log likelihood	$-1.376.10^6$	
log likelihood (L_0)	$-2.252.10^6$	

Table 2. Estimation of rounding model for departure times.

To illustrate the meaning of the estimates we compute the implications for the rounding probabilities when the actual observation is 19 minutes after the hour. The following rounding possibilities and the corresponding probabilities are found:

to 0 minutes after the hour (the nearest multiple of 60):	2.1 %
to 15 minutes after the hour (the nearest multiple of 15):	29.0 %
to 19 minutes after the hour (no rounding):	4.6 %
to 20 minutes after the hour (the nearest multiple of 5):	46.4 %
to 30 minutes after the hour (the nearest multiple of 30:	17.9 %.

The estimation result in Table 2 means that rounding to multiples of 5 minutes dominates when we consider an individual observation. Note, however, that rounding to a certain multiple of 5 (say n) only takes place for the 4 nearest neighbours (n-2, n-1, n+1, n+2). With the multiples of 15, 30 and 60 the numbers of these neighbours are 14, 29 and 59, respectively. Thus the base values for a_5 to

a_{60} have to be multiplied with factors 4 to 59 when one wants to know the total number of reported departure times. In that case the 30 minute multiple comes out as the most frequently mentioned one, and this is clearly what comes out when one considers the original data.

Estimation of model; arrival times.

A similar approach has been applied to arrival time data. The raw data are presented in Figure 2. It shows a pattern that is rather similar to the departure time figures, although the scores are less peaked in multiples of 5. The share of unrounded departure times is clearly higher (about 15 per cent is rounded to a value like 1,2,3,4,6,7, etc., as opposed to about 5 per cent for arrival times).

Figure 2. Distribution of reported arrival times.

Estimation results are shown in Table 3. The results of the arrival time estimates are to some extent similar to the departure time roundings: the 60 minute rounding is very unimportant, and the b values are negligible, except b_5. A striking difference between departure and arrival times is that rounding to a multiple of 5 is much more dominant for arrival times. As an illustration we compute again the rounding probabilities when the actual time of arrival is 19 minutes after the hour:
to 0 minutes after the hour (the nearest multiple of 60): 0.0 %
to 15 minutes after the hour (the nearest multiple of 15): 9.3 %
to 19 minutes after the hour (no rounding): 10.4 %
to 20 minutes after the hour (the nearest multiple of 5): 76.0 %
to 30 minutes after the hour (the nearest multiple of 30): 4.3 %.
Thus, rounding to multiples of 5 minutes is quite dominant. Absence of rounding is the next important one and rounding to the nearest multiple of 15 is rather unimportant. Rounding probabilities to multiples of 30 and 60 minutes are small.

coefficient	Maximum likelihood estimate	standard error
a_5	0.900	0.00201
b_5	-0.1400	0.00127
a_{15}	0.065	0.00165
b_{15}	0.0071	0.00028
a_{30}	0.014	0.00146
b_{30}	0.0026	0.00014
a_{60}	-0.00005	0.00014
b_{60}	0.00006	0.00002
log likelihood	$-1.615.10^6$	
log likelihood L_0	$-2.252.10^6$	

Table 3. Estimation of rounding model for arrival times.

Distribution of actual departure times conditional on reported departure times.

We finish this discussion with noting that we have now derived the distribution of *reported departure times, conditional on the actual departure time*. One may also be interested to derive the reverse: the distribution of the *actual departure time conditional on the reported departure time*. For example, when the reported time of departure m equals 15 minutes, what is the probability that the actual time n equals 8, 9, 10, etc. This can be achieved by using Bayes' formula (Hogg and Craig, 1970). Let $p_{m,n}$ be the probability of the reported time m given the actual departure time n (estimated above), and let g_n be the distribution of actual departure times. Then the joint density f(m,n) of m and n equals

$$f(m,n) = p_{m,n}.g_n$$

Since we want to determine k(n|m), the distribution of the probability of an actual arrival at n given a reported value m, we make use of the Bayes' formula:

$$k(n|m) = [p_{m,n}.g_n]/[\, p_{m,0}.g_0 + p_{m,1}.g_1 + ... + p_{m,59}.g_{59}].$$

Since we assume that the density of the actual departure time g(n) is given as:

$$g_n = 1/60 \text{ for } n=0,...,59,$$

Bayes' formula can be simplified as:

$$k(n|m) = p_{m,n} / [\, p_{m,0} + p_{m,1} + ... + p_{m,59}].$$

Application of this formula to for example k(4,4) implies that k(4,4) = 1: when the reported time of departure equals 4, one can be sure that the actual departure time equals 4. On the other hand, we find the following probabilities for the actual values underlying the reported observation m=15

Actual departure time n	Probability of actual departure time given reported value of departure time m=15
8	4.3
9	4.3
10	4.3
11	4.3
12	4.3
13	10.8
14	11.4
15	12.5
16	11.4
17	10.8
18	4.3
19	4.3
20	4.3
21	4.3
22	4.3

Table 4. Probability (%) of actual departure time (n=8,…,22) given a reported departure time of m=15.

The table shows that a reported departure time of m=15 means that the probability that the actual departure time is indeed 15, is only equal to 12.5%. The higher probabilities for the actual departure time are found in the range between 13 and 17 minutes, but the share for the remaining departure times further away is still substantial (43%).

Information of this type can be used in further statistical analyses of travel data in travel behaviour to give an adequate representation of errors in variables (see for example Johnston, 1984). An important implication of our approach is that rounded observations of travel times have a much larger variance than unrounded ones. For example, in our approach the reported duration of a trip being equal to 32 minutes has a much smaller variance than a trip with a reported duration of 30 minutes[4]. Such differences in variance are not well captured in standard econometric methods.

3. Discussion

One may wonder why the rounding rules applied with arrival times are more accurate than those with departure times (rounding to multiples of 15 and 30 minutes take place much less frequently). Various explanations might be thought of.

The structure of the questionnaire. The question on the times of departure and arrival are posed in an identical way: *"At what time did you depart/arrive? .... hour .... min"*. Note that these questions invite respondents to give an exact specification of the departure/arrival time. We conclude that the difference in the rounding practice for arrivals and departures cannot be explained by the way the questions are phrased.

Another point is that most respondents will fill out the questionnaire *at the end of the day*. Many of them will have forgotten the exact minutes of departure and arrival of trips made 3-15 hours earlier. This explains the practice of rounding, but it does not explain why it occurs more often with departures than with arrivals.

Structure of public transport timetables. A bias of public transport timetables towards multiples of 30 minutes as frequently used departure times might influence the reported departure times[5]. Such a timetabling practice, however, does not exist in The Netherlands. Note also that departure times reported here relate to the whole chain, so that the departure time would not indicate the time of

[4] For example, in the most extreme case a two-minute trip with a departure at 8:14 and arrival at 8:16 may be reported as a 30 minute trip after rounding. The same holds true for a 58 minute trip that started at 8:16 and ended at 9.14. This illustrates the large range on which a trip with a reported duration of 30 minutes may be based. On the other hand a trip starting at a reported time of 8.16 and ending at 8.48 will just have lasted 32 minutes according to our model, implying a zero variance (remember that apart from rounding, all other data errors are ignored in our analysis).

[5] The share of public transport in the total number of trips is about 5 per cent in the Netherlands. Its share in the total number of kilometres travelled is about 13%.

departure of the train, but the time the respondent leaves to make a trip. A final observation is that in developed countries the only collective transport mode that does not use timetables at the *one minute level of precision* is aviation (it uses multiples of 5).

As opposed to time tables of public transport, *most non-transport activities that have a scheduled character start at multiples of 15, 30 or 60 minutes:* examples are hours at school, meetings, appointments, work, church services, sport events, cinema performances, etc. In some cases both the time of the beginning and the end are exactly specified, but often the beginning is more rigid and explicit than the end. These phenomena may lead one to expect that an important share of activities of persons *start* at multiples of 15, 30 or 60 minutes and that a smaller share of these activities *end* at multiples of 15, 30 or 60 minutes. The consequence is that one expects the concentration of reported times at multiples of 15, 30 and 60 minutes to be larger for arrivals than for departures. However, the data reveal that the opposite takes place. On the other hand there are many activities that do not have a scheduled character. For example the arrival at home after an activity is usually not followed by an activity scheduled at an exact point in time. Thus the share of scheduled activities in activity patterns must not be exaggerated.

Another point is that the start/end of an activity does not necessarily coincide with the arrival/departure of a trip. In many cases there are *transitory activities* (relax, wait, talk to other participants, deposit one's coat at the cloak room, report at the entrance, find your way to the exact place of the activity, wait for the elevator, etc.). The Dutch travel survey (like many other travel surveys) does not specify these transitory activities so that it is left to the respondent whether he considers them as part of the trip, or of the activity carried out. Consider the case of a student where a lecture is scheduled to end at 12:45 sharp, in reality it ends at 12:47, the student talks to his class mates until 12:49, he leaves the university building at 12:53 to walk to the parking place of his car which he starts to drive at 12:56. Then he may answer the question *'at what time did you leave'* by filling out any of the above mentioned times, plus rounded times such as 12:45, 12:50, 12:55 and 13:00 o'clock. A similar story of course holds true for transitory activities *before* a scheduled activity.

The question remains why people are more inclined to round with departure times than with arrival times. Probably, the most important answer is that *scheduled activities force people to plan their trips in advance which provides them with anchor points for their memory afterwards.* At the end of the day they will still remember whether they arrived long before the scheduled time, or whether they were late. Since, as mentioned above, scheduling takes place more often in terms of the *start* of an activity rather than the *end*, people will have more precise memories about the time of arrival and they will therefore also have a tendency to apply rounding less frequently than with departures. This sheds some light on the literature of scheduling. As put forward for example by Small (1982, 1992) and Wilson (1989), travellers face the problem of arriving in time at scheduled activities (like the start of work, or the start of a business meeting). Given a high penalty for arriving late, travellers tend to depart at such a moment that they will arrive in time. When transport systems are unreliable (congestion caused by non-recurrent events, delays or missed connections in public transport) travellers will plan their trip in such a way that delays can be accommodated. This means that one may expect travellers to arrive early in case of scheduled activities with penalties and uncertainty in travel times. The functioning of a penalty for arriving late means that the traveller will have a keen eye on whether he really arrived early or late. When he arrives early, the traveller will have an additional type of transitory activity, compared with the ones mentioned above: safety time to avoid being late.

Thus, we arrive at several differences between the start and the end of an activity. First, the start is more often fixed in time than the end is. Second, the element of transport system uncertainty is present for the person who wants to meet the requirement of being in time at the start, it does not play a role at the end of the meeting. Third, the penalty of arriving late at the start may be perceived to be larger than the penalty of leaving early[6]. These three differences imply that on average travellers will be much more concerned about the starting time of activities than the time they end.

We finish this section with a discussion of the possible implications of two assumptions on which the above estimations are based: uniform distribution of actual departure times during an hour, and absence of measurement errors. The assumption that departure and arrival minutes are distributed uniformly was made since we have no prior knowledge about the distribution of the exact minute in the hour during which departures take place One might argue that since scheduled activities usually end at 0, 15, 30 or 45 minutes after the hour, there will be a tendency that the density of actual departure times is higher at those times. This would offer an alternative interpretation for the empirical results. With the given data this alternative interpretation cannot be falsified. However, it may be argued that it is not a very plausible explanation for several reasons. First of all we can make use of other data sources where we have both actual and reported departure time data. From a survey in the USA (Battelle, 1997) among car drivers in Lexington, it appears that the distribution of actual departure times is very close to uniform. The second reason is that transport statistics show that a considerable part of human activities are not strictly scheduled: in the Netherlands more than half of all moves relate to activities such as shopping, recreation and social visits (CBS, 1998). Therefore, an outcome that 95 per cent of the actual departures would take place at 'round' minutes (i.e., at multiples of 5) would be implausible. Another reason is that this explanation ignores the importance demonstrated above of transitory activities taking place between the end of an activity and the start of a trip. Another argument concerns trips where scheduled public transport services are used. The departure times at bus stops and railway stations tend to be distributed uniformly during the hour, so that one would expect a uniform distribution of departure times as formulated in section 2[7]. Also the discussion given above of the difference between the distribution of departure and arrival times strongly supports the view that the peaks in the distribution of reported times are due to rounding and not to peaks in actual times. We noted that if an activity is scheduled, usually the certainty about its starting point is higher than about its end point. Therefore, if the distribution of reported departure and arrival times would be dictated by the actual start of these activities, one would expect larger peaks in the distribution of arrival times compared with departure times, but in reality the opposite is true[8].

[6] We do not go into details about chaining activities with fixed start and end times. When there is sufficient travel time between the two, this leads to an additional type of transitory activity. When the time is not long enough, the traveller by his temporal behaviour reveals which of the two activities will have the higher penalty (leaving early versus arriving late).

[7] What really matters is not the official departure time of the public transport services, but the departure time of the traveller from his origin, thus taking into account the access time to the public transport node. Thus, even should there be a tendency for public transport time tables to be biased towards departure times of the services that are multiples of 5 minutes, the variance in the access times would make this invisible when departure times of travellers are considered.

[8] Another possibility with arrival times is that the distribution of actual times has high probabilities at times just before 'round' minutes because most people try to arrive in time. However, inspection of the reported arrival times

We conclude that with the given data we cannot test whether the distribution of actual departure minutes is uniform. It is highly implausible, however, that a non-uniform distribution is the sole reason for the peaks in the reported departure times. One cannot exclude, however, the possibility that there is a tendency that densities of arrival and departure times are higher at ‘round’ minutes than around other minutes. As far as this is true, it would imply that we have overestimated the rounding tendency. Given the above arguments, this tendency towards overestimation is most probably small.

The second assumption in section 2 that may need some discussion concerns the premise that rounding is the only source of errors when reporting departure and arrival times: it is ruled out in the statistical analysis that people report wrong departure times because of mistakes, inaccurate watches or bad memory. Of course such errors will take place frequently in travel surveys and they will partly express themselves in rounding. For example, a bad memory of the exact times at the end of the day will induce many respondents to use proxies. In those cases where these mechanisms do not express themselves via rounding, they contribute to the variance of error in observed data, but there is no reason to expect that they will lead to systematic distortions in the analysis of rounding[9].

4. Concluding remarks.

Our analysis of departure and arrival times indicates that rounding is a rule, rather than an exception. About 5-15 per cent of all reported times assume values that are not multiples of 5, whereas these are 80 per cent of the possible clock times. In the case of scheduled activities the reported times are probably more precise since scheduling implies the use of anchor points in the time frame. With fixed schedules there may be a high penalty for being late so that travellers will be induced to keep record of the exact timing of the trips. Since scheduling of start times takes place more often than of end times, it is plausible that reported times of arrival are more accurate than reported times of departure.

In the research on travel behaviour, data on travel times usually play an important role. These travel times follow as the result of subtracting reported times of arrival and departure. Given the large rounding errors observed here, it is clear that errors in reported travel times (and related variables such as travel speeds) will be large. This ‘error in data’ phenomenon will obviously hamper the analysis of data on individual travel behaviour. In the present paper we have developed a method, based on a Bayesian approach to derive the probability that a reported arrival time m means that the actual arrival time equals n. This method can be used in ‘errors in variable methods’ to give an adequate representation of the measurement error. We demonstrated that the variance of rounded travel times is much larger than that of unrounded ones. This approach must be considered superior to the usual approach where all measurement error is supposed to be represented by a common variance.

---

does not reveal such a tendency. For example the data in Figure 2 even demonstrate a slight tendency in the opposite direction: the share of respondents reporting they arrived between 1-15 minutes after the hour is somewhat larger than the share reporting they arrived between 45-59 minutes (26 versus 22 per cent).

[9] Note also that without additional data, adding an error term e_m with mean zero and variance s to the model such that the reported departure time is equal to the actual departure time plus e_m will not yield meaningful estimates of s.

Rounding has a larger impact than just affecting the variance of travel times, however. Given the large scale at which rounding takes place, it may also affect averages computed on the basis of national surveys when probabilities of rounding upward and downward do not cancel. Consider for example the distribution of reported trip duration in The Netherlands. This distribution is skewed: the most frequently reported trip duration (mode) is 10 minutes, the median value is 15 minutes, and the mean value is about 20 minutes. Therefore, the number of trips with an actual duration of between 15 and 30 minutes will be considerably larger than the number of trips with a duration between 30 and 45 minutes. As a result, the probability of rounding upward is considerably higher than the probability of rounding downward[10]. The conclusion is that in this case rounding of arrival and departure times leads to overestimates of average travel times[11].

Finally, ignoring the rounding problem could lead to erratic patterns when the travel survey data would be used to give a minute to minute record of the number of travellers in the transport system. Consider for example the 24 hour average number of people in transport in each minute for our sample of 550,000 respondents. The departure and arrival data indicate that during the first minute of the hour 120,000 persons enter the transport system, whereas only 55,000 persons leave. This would imply a sudden net increase of 65,000 persons during one minute, which is much higher than the small net decreases during subsequent minutes of about 1500 persons per minute. This obviously hinders a proper assessment of the development of the number of persons in traffic in the course of time. By using the Bayesian approach presented in section 2, this problem can be overcome.

In our discussion of rounding we touched on the importance of *transitory activities* in scheduled activity patterns. These transitory activities are often ignored in the analysis of travel behaviour. A main reason for these transitory activities is that they emerge in a response to reduce the penalty of arriving late at a scheduled activity. They also result from low frequency services in public transport. Transitory activities are important to reduce bottlenecks in internal and external transport systems. An example of internal transport systems is that the elevator capacity usually will not allow everybody to arrive just in time or leave immediately after, at larger events. Similarly, parking facilities do not function well under these circumstances. An example of external transport systems concerns the capacity to absorb visitors for large-scale events in stadiums, exhibition centres, etc. Transitory activities do not only keep bottleneck problems manageable, they may also have value per se for the travellers. They deserve more attention in transport behaviour than they usually get. For a proper analysis of their presence and size rather detailed questionnaires are needed.

A final point of attention is the possibility of linking reported travel time data to archived global positional data. The combination of GIS and GPS offers substantial potential to improve the quality of data on travel time and travel distance in passenger surveys. This does not only hold true for automobile trips but probably also for other kinds of trips (Quiraga and Bullock, 1998, Uchida et al, 2001).

---

[10] This implies that the figures of 20 and 15 minutes mentioned in the text for mean and median are biased. The effect on the mean is probably larger than on the median.

[11] In the study of Battelle (1997) a comparison of reported and actual travel times indeed reveals that reported travel times based on recall generally overstate travel time. A similar conclusion was drawn about travel distances.

Acknowledgement.
The author thanks Uty Pang Atjok who provided computational assistance. In addition David Greene and two anonymous referees gave constructive comments.

References:

Battelle Transportation Division, 1997. Lexington Area Travel Data Collection Test, Global Positioning Systems for Personal Travel Surveys, Columbus, Ohio.

CBS (Central Bureau of Statistics), 1998. *De mobiliteit van de Nederlandse bevolking (The mobility of the Dutch population),* Voorburg.

CBS (Central Bureau of Statistics), 1996. *Auto's in Nederland (Cars in the Netherlands),* Kluwer, Deventer.

Hogg, R.V. and A.T. Craig, 1970. *Introduction to Mathematical Statistics,* Macmillan, London.

Johnston, R.J., 1984. *Econometric Methods*, McGraw-Hill, New York.

Levine, R.V., 1997. *A Geography of Time,* Basic Books, New York.

Quiroga, C.A. and D. Bullock, 1998. Travel time studies with global positioning and geographic information systems, an integrated methodology, *Transportation Research C*, vol. 6, pp. 101-127

Rietveld, P., B. Zwart, B. van Wee, T. van den Hoorn, 1999. On the relationship between travel time and travel distance of commuters: reported versus network travel data in the Netherlands, *The Annals of Regional Science*, vol. 33, pp. 269-288.

Small, K.A., 1982. The scheduling of consumer activities: work trips, *American Economic Review*, vol. 72, pp.467-479.

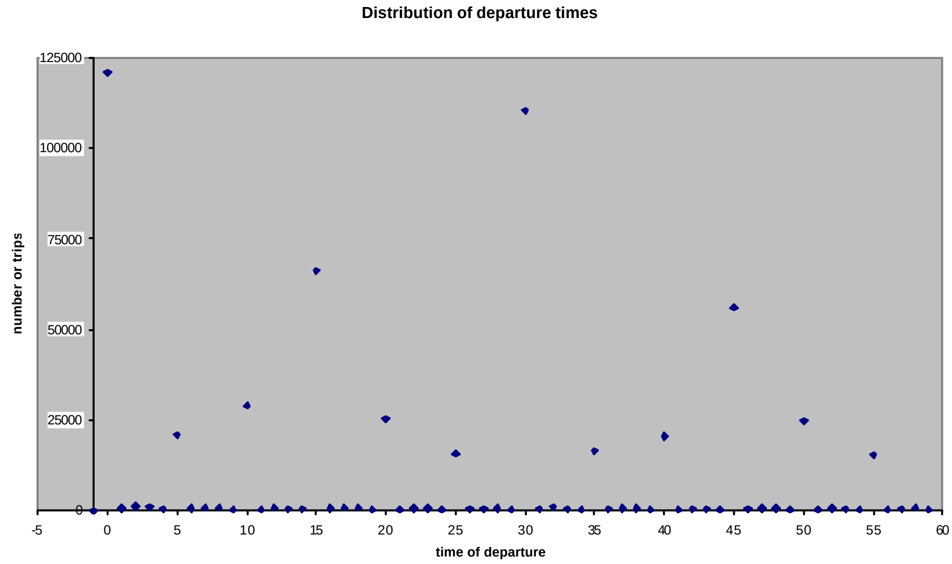
Small, K.A., 1992. *Urban Transportation Economics*, Harwood Publ., Luxemburg.

Uchida, T., M. Pursula, A. Suzuki, Y.H. Lee, D. Takehiko, 2001. Monitoring personal movement: development of PEAMON (Personal Activity Monitor) for automated trip-diary, Tohoku University Japan.

US Department of Transportation, Bureau of Transportation Statistics (BTS), 1997. *Nationwide Personal Transportation Survey*, Washington, DC.

Wilson, P.W., 1989. Scheduling costs and the value of travel time, *Urban Studies*, vol. 26, 356-366.

Zahavi, Y., 1977. *The UMOT model*, World Bank, Washington.


Distribution of departure times
number or trips
time of departure
125000
100000
75000
50000
25000
0
-5
0
5
10
15
20
25
30
35
40
45
50
55
60

Distribution of arrival times